The Structure of Geographical Threshold Graphs

Milan Bradonjić[1], Aric Hagberg[2], Allon G. Percus[3]

[1] Department of Electrical Engineering, UCLA, Los Angeles, CA 90095
and Mathematical Modeling and Analysis Group, Theoretical Division,
Los Alamos National Laboratory, Los Alamos, NM 87545, milan@ee.ucla.edu
[2] Mathematical Modeling and Analysis Group, Theoretical Division,
Los Alamos National Laboratory, Los Alamos, NM 87545, hagberg@lanl.gov
[3] Information Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545
and Department of Mathematics, UCLA, Los Angeles, CA 90095,
percus@ipam.ucla.edu

Abstract. We analyze the structure of random graphs generated by the geographical threshold model. The model is a generalization of random geometric graphs. Nodes are distributed in space, and edges are assigned according to a threshold function involving the distance between nodes as well as randomly chosen node weights. We show how the degree distribution, percolation and connectivity transitions, clustering coefficient and diameter relate to the threshold value and weight distribution. We give bounds on the threshold value guaranteeing the existence and absence of a giant component, connectivity and disconnectivity of the graph, and small diameter. Finally, we consider the clustering coefficient for nodes with a given degree l, finding that its scaling is very close to $1/l$ when the node weights are exponentially distributed.

Key words: random graph, geographical threshold graph, giant component, connectivity, clustering coefficient.

1 Introduction

Large networks such as the Internet, World Wide Web, phone call graphs, infections disease contacts, and financial transactions have provided new challenges for modeling and analysis [Bon05]. For example, Web graphs may have billions of nodes and edges, which implies that the analysis on these graphs, i.e., processing and extracting information on these large sets of data, is “hard” [APR02]. Extensive theoretical and experimental research has been done in web-graph modeling. Early measurements suggested that the Internet exhibits a power-law degree distribution [FFF99] and that the web graph also follows a power-law distribution in in- and out-degree of links [KKR$^+$99]. Modeling approaches using random graphs have attempted to capture both the structure and dynamics of the web graph [KRR$^+$00,BA99,ACL00,BRST01,CF01].

The study of random graphs began with the introduction of the uniform random graph model [ER59,ER60]. Since then many other models have been proposed to better capture the structure seen in real-world networks [Bol01,Dur06]. Some examples are random graph models with a given or expected degree sequence [MR95,CL06], threshold graphs [MP95,HSS06] with edges created according to a function of node weights, or graphs with an underlying geometric structure, such as random geometric graphs [Pen03]. In this paper we study another recent addition to this collection of models: geographical threshold graphs (GTGs), a static model for networks that includes both geometric information and node weight information.

GTGs combine the geometric structure of random geometric graphs with node properties similar to threshold graphs. The properties of this graph ensemble have only recently begun to be studied [MMK05,BHP07,BK07]. One motivation for analyzing this model is that many real networks need to be studied by using a "richer" stochastic model than random geometric graphs. The GTG model has been applied, for instance, in the study of wireless ad hoc networks in systems where the wireless nodes have different communication ranges or battery power [BK07]. In that case, the weights represent available power or bandwidth of a wireless node. By varying the weights, properties such as the diameter or degree distribution can be tuned. Other possible applications of GTGs that are yet to be explored are epidemic modeling, where the weights could represent susceptibility or infectivity of an individual, or social networks where the weights might be related to affinity or attractiveness.

2 Geographical Threshold Graph Model

The GTG model is constructed from of a set of n nodes placed independently and uniformly at random in a volume in $\mathbf{R}^d$. A non-negative weight w_i, taken randomly and independently from a probability distribution function $f(w) : \mathbf{R}_0^+ \to \mathbf{R}_0^+$, is assigned to each node v_i for $i \in [n]$. Let $F(x) = \int_0^x f(w)dw$ be the cumulative density function. For two nodes v_i and v_j at distance r, the edge (i,j) exists if and only if the following connectivity relation is satisfied:

$$G(w_i, w_j)h(r) \geq \theta_n\,, \tag{1}$$

where θ_n is a given threshold parameter that depends on the size of the network. The function $h(r)$ specifies the connection probability as a function of distance and is assumed to be decreasing in r. In the following we take $h(r) = r^{-\beta}$, for some positive β, which is typical for e.g., the path-loss model in wireless networks [BK07]. The interaction strength between nodes $G(w_i, w_j)$ is typically taken to be symmetric (to produce an undirected graph) and either multiplicatively or additively separable, i.e., in the form of $G(w_i, w_j) = g(w_i)g(w_j)$ or $G(w_i, w_j) = g(w_i) + g(w_j)$.

Some basic results have already been shown. For the case of uniformly distributed nodes over a unit space it has been shown [MMK05,BK07] that the expected degree of a node with weight w is

$$\mathrm{E}[k(w)] = \frac{n\pi^{d/2}}{\Gamma(d/2+1)} \int_{w'} f(w')\big(h^{-1}(\theta_n/G(w,w'))\big)^d dw', \tag{2}$$

where h^{-1} is the inverse of h. The degree distribution has been studied for specific weight distribution functions $f(w)$ [MMK05]. In both the multiplicative and additive case of $G(w,w')$, questions of diameter, connectivity, and topology control have been addressed [BK07].

Here we restrict ourselves to the case of $g(w) = w$, $\beta = 2$, and nodes distributed uniformly over a two-dimensional space. For analytical simplicity we take the space to be a unit torus, and use the additive model for the connectivity relation.

$$\frac{w_i + w_j}{r^2} \geq \theta_n. \tag{3}$$

Certain of our techniques may be generalized to other cases in a straightforward manner. Finally, we impose the following relatively weak conditions on the weight distribution $f(w)$: (1) a finite mean $\mu = \mathrm{E}[w]$ and (2) a finite variance $\sigma^2 = \mathrm{E}[w^2] - \mathrm{E}[w]^2$. Some examples of GTG instances with exponential weight distribution $f(w) = e^{-w}$ are shown in Figure 1.

The paper is organized as follows. We first state a basic property concerning the degree distribution of GTGs. In Section 4, Theorems 1 and 2 provide bounds on θ_n for the absence and the existence of a giant component. Similarly, in Section 5, Theorems 3 and 4 provide bounds on θ_n for the graph being disconnected and connected. Section 6 gives upper bounds on the diameter, along with simulation results. Finally, in Section 7 we calculate the clustering coefficient, and discuss certain of its properties.

3 Degree Distribution

We start by stating the degree distribution in our GTG model. Let the position vector of the nodes be $\mathbf{x}$ and the weight vector be $\mathbf{w}$. W.l.o.g. let us consider node v_1. It is straightforward to show that the probability of v_1 having degree k, given weights $\mathbf{w}$, is

$$\Pr[d_1 = k|\mathbf{w}] = \binom{n-1}{k} \prod_{i=2}^{k+1} \mathrm{Area}(B(x_i, r_{i1})) \prod_{j=k+2}^{n} (1 - \mathrm{Area}(B(x_j, r_{j1}))), \tag{4}$$

where $\mathrm{Area}(B(x_i, r_{i1}))$ is the area of the ball at center x_i with radius r_{i1}, and due to (3) the radii are given by

$$r_{i1} = \sqrt{\frac{w_1 + w_i}{\theta_n}} \tag{5}$$

for $i = 2, \ldots, n$. After marginalization, it follows

$$\begin{aligned}
\Pr[d_1 = k|w_1] &= \Big(\prod_{i=2}^{n} \int_{w_i} f(w_i) dw_i \Big) \Pr[d_1 = k|\mathbf{w}] \\
&= \binom{n-1}{k} \Big(\int_w f(w) \frac{\pi(w_1 + w)}{\theta_n} dw \Big)^k \Big(1 - \int_w f(w) \frac{\pi(w_1 + w)}{\theta_n} dw \Big)^{n-1-k} \\
&= \binom{n-1}{k} \Big(\frac{\pi(w_1 + \mu)}{\theta_n} \Big)^k \Big(1 - \frac{\pi(w_1 + \mu)}{\theta_n} \Big)^{n-1-k}.
\end{aligned} \tag{6}$$

That is, the degree distribution d_i, of a node v_i with weight w_i, follows the Binomial distribution

$$d_i(\cdot|w_i) \sim Bin(n-1, p_i) \tag{7}$$

where

$$p_i = \frac{\pi}{\theta_n}(w_i + \mu). \tag{8}$$

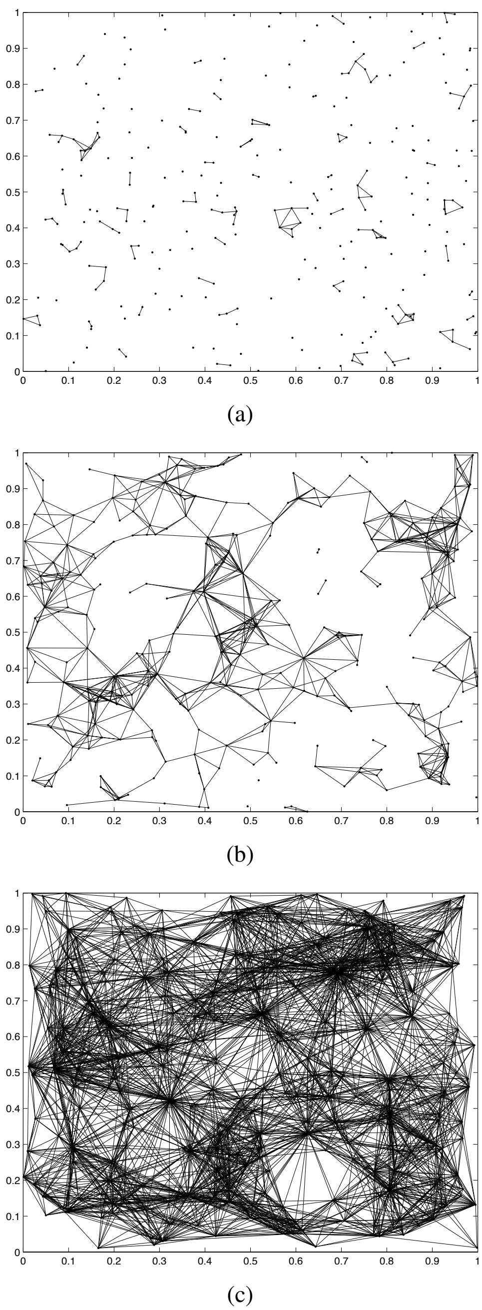

Fig. 1. Instances of GTG with exponential weight distribution, for $n = 300$ at decreasing threshold parameter values (increasing mean degree): (a) $\theta_n/n = 2\pi$, well below the percolation transition; (b) $\theta_n/n = 1$, above the percolation but below the connectivity transition; (c) $\theta_n/n = 1/2e$, well above connectivity.

4 Giant Component

Definition 1 (Giant Component). *The giant component is a connected component with size* $\Theta(n)$.

In this section we analyze the conditions for the existence of the giant component, giving bounds on the threshold parameter value θ_n where it first appears. For $\theta_n = cn$, we specify positive constants $c' > c''$ and prove that whp (with probability $1 - o(1)$), if $c > c'$ the giant component does not exist whereas if $c < c''$ the giant component exists.

We do not prove that there is a zero-one law governing the emergence of the giant component. However, given that the the probability of a giant component is 0 above c' and 1 below c'', it would be rather surprising if the transition were not a sharp one.

4.1 Absence of Giant Component

Theorem 1. *Let* $\theta_n = cn$ *for* $c > c'$*, where* $c' = 2\pi\mu$*. Then whp there is no giant component in GTG.*

Proof. We start by introducing a slightly different GTG model from our usual one. Consider the space of possible node positions and weights, $S = \{(x,y,w) : x,y \in [0,1], w \geq 0\}$. In the model we have already defined, we place n nodes in S, leading to a binomial degree distribution: call this the *Binomial GTG*. Let us now instead place nodes in S according to a spatial Poisson process with rate $nf(w)$, so that the expected number of nodes is n: call this the *Poisson GTG*. We will prove that the Poisson GTG does *not* have a giant component. It is straightforward to see that if the Binomial GTG had a giant component with nonvanishing probability, the analogous Poisson GTG would as well. Thus, the Binomial GTG cannot have a giant component either.

The proof's approach is similar to one given in [AS00]. Divide the nodes into three classes: alive, dead and neutral. Denote the number of alive nodes as Y_i. Now apply the following algorithm. At time $t = 0$, designate one node (picked u.a.r.) as being alive and all others as neutral. At each subsequent time step t, pick a node v_t u.a.r. from among those that are alive, and then consider all neutral nodes connected to v_t. Denote the number of these nodes as Z_t. Change these nodes from neutral to alive, and change v_t itself from alive to dead. The random variables Y_i, Z_i satisfy the following recursion relation: $Y_0 = 1$ and $Y_t = Y_{t-1} + Z_t - 1$, for $t \geq 1$. The number of alive nodes satisfies

$$Y_t - 1 = \sum_{i=1}^{t} Z_i - t. \tag{9}$$

Since neutral nodes are by definition those that have not yet been explored by the algorithm, the Z_i are independent random variables. We formalize this argument as follows. For a node $v_i = (x_i, y_i, w_i)$, define $S_i \subseteq S$ as the subspace of all positions and weights of nodes that can be connected to v_i, namely $S_i = \{(x,y,w) : x,y \in [0,1], w \geq 0, (x-x_i)^2 + (y-y_i)^2 \leq (w+w_i)/\theta_n\}$. At time $t = 0$, any node within S_0 is a neutral node connected to v_0. But at a subsequent time step t, nodes within any S_i for $i < t$ have already been designated alive, so only those in

$$B_t = S_t \setminus \bigcup_{i=0}^{t-1} S_i$$

can be neutral nodes connected to v_t. Thus, the nodes figuring within Z_t and $Z_{t'}$, for any two different time steps t and t', are drawn from *disjoint* subspaces B_t and $B_{t'}$. The Z_i are simply restrictions of the Poisson process to the B_i. Due to the memoryless nature of the Poisson process, they are independent Poisson random variables. Given that $B_i \subseteq S_i$ and the expected population of S_i is np_i with p_i as defined in Eq. (8), Z_i satisfies the stochastic bound

$$\Pr[Z_i \geq k] \leq \Pr[Po(np_i) \geq k]. \tag{10}$$

Now consider nodes that are alive, and let T be the largest t such that $Y_t > 0$. Then T is the size of the component containing v_0, and the giant component exists if and only if $T = \Theta(n)$ with some nonvanishing probability. The variable T satisfies

$$\Pr[T \geq t] = \Pr[Y_t > 0] = \Pr[Y_t \geq 1] = \Pr[\sum_{i=1}^{t} Z_i \geq t] \leq \Pr[\sum_{i=1}^{t} Po(np_i) \geq t]. \tag{11}$$

We take the threshold to be $\theta_n = cn$. Since the sum of independent Poisson random variables is itself Poisson distributed, we need to prove that $\Pr[Po(n\sum_{i=1}^{t} p_i) \geq t] \to 0$ for $t = \Theta(n)$, for some $c > 0$. For any constant $\varepsilon \in (0,1)$, the following inequality holds:

$$\Pr\Big[Po(n\sum_{i=1}^{t} p_i) \geq t\Big] \leq \Pr\Big[Po(n\sum_{i=1}^{t} p_i) \geq t \Big| \sum_{i=1}^{t} w_i \in (1\pm\varepsilon)t\mu\Big] + \Pr\Big[\sum_{i=1}^{t} w_i \notin (1\pm\varepsilon)t\mu\Big]. \tag{12}$$

We will bound the first right-hand term using the concentration of Poisson random variables [Pen03]. To maximize the conditional probability, set $\sum w_i = (1+\varepsilon)t\mu$, and then let $\lambda = n\sum_{i=1}^{t} p_i = (2+\varepsilon)at\mu$ where $a = n\pi/\theta_n = \pi/c$. Now, given any constant $\gamma \in (0,1)$, for $t \to \infty$, i.e., $\lambda \to +\infty$, it follows that

$$\Pr[Po(\lambda) \notin (1\pm\gamma)\lambda] \leq e^{-\lambda H(1-\gamma)} + e^{-\lambda H(1+\gamma)} \to 0, \tag{13}$$

where the function $H(x) = 1 - x + x\log x$, for $x > 0$. It is now sufficient to choose c large enough that $t > (1+\gamma)\lambda$, which occurs when $c > (2+\varepsilon)(1+\gamma)\pi\mu$. It follows that for any $c > 2\pi\mu$, ε and γ can be set so that the first right-hand term in Eq. (12) goes to zero.

Now consider the second right-hand term. By the central limit theorem, $(\sum w_i - t\mu)/(\sqrt{t}\sigma)$ tends to the normal distribution $N(0,1)$ as $t \to \infty$, so

$$\Pr\Big[\sum_{i=1}^{t} w_i \notin (1\pm\varepsilon)t\mu\Big] = \Pr\Big[\frac{\sum w_i - t\mu}{\sqrt{t}\sigma} \notin (-\varepsilon,\varepsilon)\sqrt{t}\frac{\mu}{\sigma}\Big] \to 0 \tag{14}$$

for any constant ε.

Thus, for $c > 2\pi\mu$, the probability that $T = \Theta(n)$ goes to zero, and so there is no giant component. $\square$

4.2 Existence of Giant Component

Theorem 2. *Let $\theta_n = cn$ for $c < c'' = \sup_{\alpha\in(0,1)} \alpha F^{-1}(1-\alpha)/\lambda_c$, where $\pi\lambda_c \approx 4.52$ is the mean degree at which the giant component first appears in Random Geometric Graphs (RGG) [Pen03]. Then whp the giant component exists in GTG.*

Proof. For any constant $\alpha \in (0,1)$, we prove that whp there are αn "high-weighted" nodes, all with weights greater than or equal to some s_n; we state s_n later. Let X_i be the indicator of the event $w_i \geq s_n$. Then $\Pr[X_i = 1] = 1 - F(s_n) =: q$. Let $X = \sum_{i=1}^{n} X_i$ be the number of high-weighted nodes. Using the Chernoff bound $\Pr[X \leq (1-\delta)\mathrm{E}[X]] \leq \exp(-\mathrm{E}[X]\delta^2/2)$, with $\delta = 1 - \alpha/q$,

$$\Pr[X \leq \alpha n] = \Pr[X \leq (1-\delta)\mathrm{E}[X]] \leq \exp\big(-n(q-\alpha)^2/(2q)\big) = n^{-\gamma} \tag{15}$$

for some constant $\gamma > 1$ satisfying $(q-\alpha)^2 = 2q\gamma \log n/n$. Solving that quadratic equation in q gives $q = \alpha + \Theta(\log n/n)$, so $F(s_n) = 1 - q = 1 - \alpha - \Theta(\log n/n)$. For any $\varepsilon > 0$ and n sufficiently large the following is satisfied

$$F^{-1}(1-\alpha) \geq s_n \geq F^{-1}(1-\alpha-\varepsilon). \tag{16}$$

Thus, let us define the sequence s_n by its limit

$$s_n \to F^{-1}(1-\alpha) = \Theta(1). \tag{17}$$

Now we consider the set of αn high-weighted nodes. For each such node v_i with weight w_i, define its characteristic radius to be

$$r_t^2(w_i) = w_i/\theta_n. \tag{18}$$

Then it follows that any other high-weighted node v_j within this radius is connected to v_i, since the connectivity relation is satisfied:

$$(w_i + w_j)/r^2 \geq w_i/r_t^2 = \theta_n. \tag{19}$$

Let $\theta_n = cn$, where $c < \alpha F^{-1}(1-\alpha)/\lambda_c$. For the radius r_t, whp it follows

$$r_t^2(w_i) = \frac{w_i}{\theta_n} \geq \frac{s_n}{\theta_n} > \frac{\lambda_c}{\alpha n}. \tag{20}$$

Let us therefore consider small circles, with a fixed radius r_0 s.t. $\sqrt{s_n/\theta_n} > r_0 > \sqrt{\lambda_c/(\alpha n)}$, around each of these αn nodes. A subgraph of this must be a RGG with mean degree $> \pi\lambda_c$, which whp contains a giant component. Since its size is $\Theta(\alpha n) = \theta_n$, it is a giant component of the GTG too. We may optimize the bound by taking the supremum of c over $\alpha \in (0,1)$, and the theorem follows. □

4.3 Comparison of Upper and Lower Bounds

We again stress that we have not proven a zero-one law for the emergence of the giant component. If a sharp transition does indeed exist, c' and c'' provide bounds on its location. Here we consider the size of the gap between the two bounds.

Claim. For any weight distribution $f(w)$, $c'/c'' \geq 2\pi\lambda_c \approx 9.04$.

Proof. First consider $c' = 2\pi\mu$. Using the telescope formula, μ satisfies

$$\mu = \int_0^\infty (1 - F(y))dy. \tag{21}$$

Now consider $c'' = \sup_{\alpha\in(0,1)} \alpha F^{-1}(1-\alpha)/\lambda_c$. We have $F : [0,+\infty) \to [0,1)$. Since F is a bijection, the inverse $F^{-1} : [0,1] \to [0,+\infty)$ exists. Let $x = F^{-1}(1-\alpha)$, and consequently $\alpha = 1 - F(x)$. Then

$$\sup_{\alpha\in(0,1)} \alpha F^{-1}(1-\alpha) = \sup_{x\in(0,\infty)} x(1-F(x)). \tag{22}$$

Define the function

$$g(x) = \int_0^x (1-F(y))dy - x(1-F(x)). \tag{23}$$

Since $g'(x) = xf(x) \geq 0$ and $g(0) = 0$, we know that $g(x) \geq 0$ for every $x \geq 0$. Let x_0 be the value at which $x(1-F(x))$ has its supremum. Then,

$$\mu - \sup_{\alpha\in(0,1)} \alpha F^{-1}(1-\alpha) = \int_0^\infty (1-F(y))dy - x_0(1-F(x_0)) \geq g(x_0) \geq 0, \tag{24}$$

from which the claim follows. □

Remark 1. For the exponential distribution $f(w) = \gamma\exp(-\gamma w)$, $c' = 2\pi/\gamma$.

Remark 2. If $\alpha F^{-1}(1-\alpha)$ has an extremum for $\alpha \in (0,1)$, this occurs at

$$\alpha = F^{-1}(1-\alpha) f(F^{-1}(1-\alpha)). \tag{25}$$

For example, for the exponential distribution the maximum is at $\alpha = 1/e$, giving a bound of $c'' = 1/e\gamma\lambda_c$.

Remark 3. Given the recent bound [KY06] $\lambda_c \geq 4/(3\sqrt{3})$ for RGG, $c'' \leq \sup_{\alpha\in(0,1)} \alpha F^{-1}(1-\alpha)3\sqrt{3}/4$. For the exponential distribution this gives a bound of $c'' \leq 3\sqrt{3}/4e\gamma$.

5 Connectivity

Definition 2 (Connectivity). *The graph on n vertices is connected if the largest component has size n.*

In this section we analyze conditions for connectivity, giving bounds on the threshold parameter θ_n at which the entire graph first becomes connected. Similarly to our approach in the case of the giant component, for $\theta_n = cn/\log n$, we specify positive constants $c' > c''$ and prove that whp, if $c > c'$ the graph is disconnected whereas if $c < c''$ the graph is connected.

As in the case of the emergence of the giant component, it seems likely but has not been proven that there is a sharp phase transition at which GTGs become connected.

5.1 Disconnected Graph

Theorem 3. *Let $\theta_n = cn/\log n$, where $c > \pi\mu$. Then the GTG is disconnected whp.*

Proof. For a node v_i, let Y_i be the indicator of the event that v_i is isolated. We will consider the sum

$$Y = \sum_{i=1}^{n} Y_i \tag{26}$$

and show that $\Pr[Y = 0] \to 0$. It will then follow that whp there is at least one isolated node and so the graph is disconnected.

From the binomial degree distribution in Eq. (6), the probability that v_i is isolated, conditional on w_i, is

$$\Pr[Y_i = 1 | w_i] = \left(1 - \frac{w_i + \mu}{\theta_n}\pi\right)^{n-1}. \tag{27}$$

Now define

$$\begin{aligned} p \equiv \mathrm{E}[Y_i] &= \Pr[Y_i = 1] \\ &= \int f(w_i)\left(1 - \frac{w_i + \mu}{\theta_n}\pi\right)^{n-1} dw_i. \end{aligned} \tag{28}$$

Applying the second moment method,

$$\begin{aligned} \Pr[Y = 0] &\leq \frac{\mathrm{Var}[Y]}{\mathrm{E}[Y]^2} \\ &= \frac{\sum_i \mathrm{Var}[Y_i] + \sum_{i \neq j} \mathrm{Cov}[Y_i, Y_j]}{(np)^2}. \end{aligned} \tag{29}$$

The variance and covariance are given by

$$\mathrm{Var}[Y_i] = \mathrm{E}[Y_i^2] - \mathrm{E}[Y_i]^2 = p - p^2 \tag{30}$$

$$\mathrm{Cov}[Y_i, Y_j] = \mathrm{E}[Y_i, Y_j] - \mathrm{E}[Y_i]\mathrm{E}[Y_j] = \Pr[Y_i = 1, Y_j = 1] - p^2, \tag{31}$$

so

$$\begin{aligned} \Pr[Y = 0] &\leq \frac{n(p - p^2) + n(n-1)(\Pr[Y_i = 1, Y_j = 1] - p^2)}{(np)^2} \\ &< \frac{1}{np} + \frac{\Pr[Y_i = 1, Y_j = 1]}{p^2} - 1. \end{aligned} \tag{32}$$

Let us first consider the $1/(np)$ term. Let $\theta_n = c\frac{n}{\log n}$, where c is a constant. We claim that if $c > \pi\mu$, then $1/(np) \to 0$. To see this, let ζ be any positive constant that satisfies $c > \pi(\mu + \zeta)$. We have

$$\begin{aligned} p &= \int f(w)\left(1 - \frac{w + \mu}{\theta_n}\pi\right)^{n-1} dw \\ &\geq \int_0^{\zeta} f(w)\left(1 - \frac{w + \mu}{\theta_n}\pi\right)^{n-1} dw \end{aligned}$$

$$
\begin{aligned}
&\geq F(\zeta)\Big(1-\frac{\mu+\zeta}{\theta_n}\pi\Big)^{n-1}\\
&= F(\zeta)\Big(1-\frac{\mu+\zeta}{cn}\pi\log n\Big)^{n-1}\\
&= F(\zeta)n^{-(\mu+\zeta)\pi/c}(1+o(1)).
\end{aligned}
$$

Therefore, if $c > \pi(\mu+\zeta)$, $pn \geq F(\zeta)\omega(n)$ and so $1/(np) \to 0$ for $n \to \infty$.

Next, we will show that $\Pr[Y_i = 1, Y_j = 1]/p^2 = o(1)$. Consider the joint probability conditional on a set of weights $\mathbf{w}$. Denoting the neighborhood relation by $v_i \sim v_j$,

$$
\begin{aligned}
\Pr[Y_i = 1, Y_j = 1|\mathbf{w}] &= \Pr[v_i \nsim v_j, \bigcap_{k\neq i,j} v_i \nsim v_k, v_j \nsim v_k|\mathbf{w}]\\
&= \Pr[v_i \nsim v_j|w_i, w_j]\Pr[\bigcap_{k\neq i,j} v_i \nsim v_k, v_j \nsim v_k|v_i \nsim v_j, \mathbf{w}]\\
&= \Pr[v_i \nsim v_j|w_i, w_j]\prod_{k\neq i,j}\Pr[v_i \nsim v_k, v_j \nsim v_k|v_i \nsim v_j, \mathbf{w}]. \quad (33)
\end{aligned}
$$

We now use the easily verified property that given events Q, R and S that depend on $\mathbf{w}$,

$$
\Pr[R^c, S^c|Q^c, \mathbf{w}] = 1 - \Pr[R|\mathbf{w}] - \Pr[S|\mathbf{w}] + (1 - \Pr[Q|R, S, \mathbf{w}])\frac{\Pr[R, S|\mathbf{w}]}{\Pr[Q^c|\mathbf{w}]}. \quad (34)
$$

Let $a = \Pr[v_i \sim v_j|w_i, w_j]$, $b = \Pr[v_i \sim v_k|w_i, w_k]$, $c = \Pr[v_j \sim v_k|w_j, w_k]$ and define the *clustering coefficient*

$$
C = \Pr[v_i \sim v_j|v_i \sim v_k, v_j \sim v_k, w_i, w_j, w_k]. \quad (35)
$$

Then,

$$
\Pr[Y_i = 1, Y_j = 1|\mathbf{w}] = (1-a)\prod_{k\neq i,j}\big[1 - b - c + (1-C)\frac{bc}{1-a}\big]. \quad (36)
$$

Note that $a = (w_i + w_j)\pi/\theta_n$, and similarly for b and c.

In Section 7.1, we show (Lemma 2) that if $w_i, w_j, w_k \leq \hat{w} = (1 - 3\sqrt{3}/4\pi)\theta_n/2\pi$, then $C \geq a$. Thus, under those conditions,

$$
\begin{aligned}
1 - b - c + (1-C)\frac{bc}{1-a} &\leq 1 - b - c + bc\\
&= (1-b)(1-c). \quad (37)
\end{aligned}
$$

Now average Eq. (36) over all weights. It follows from the finite variance of $f(w)$ that for any constant M, $F(M\theta_n) = 1 - o(1/n)$, and so

$$
\begin{aligned}
\Pr[Y_i = 1, Y_j = 1] &= \int f(w_i)dw_i \int f(w_j)dw_j(1-a)\Big(\int f(w_k)dw_k\Big[1 - b - c + (1-C)\frac{bc}{1-a}\Big]\Big)^{n-2}\\
&= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j(1-a)\Big(\int_0^{\hat{w}} f(w_k)dw_k\Big[1 - b - c + (1-C)\frac{bc}{1-a}\Big]\Big)^{n-2}(1+o(1))\\
&\leq \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j(1-a)\Big(\int_0^{\hat{w}} f(w_k)dw_k(1-b)(1-c)\Big)^{n-2}(1+o(1))
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j \Big\{ (1 - \frac{\pi}{\theta_n}(w_i + w_j)) \times \\
&\quad \Big(\int_0^{\hat{w}} f(w_k)dw_k (1 - \frac{\pi}{\theta_n}(w_i + w_k))(1 - \frac{\pi}{\theta_n}(w_j + w_k)) \Big)^{n-2} \Big\} \\
&= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j \Big\{ (1 - \frac{\pi}{\theta_n}(w_i + w_j)) \times \\
&\quad \Big(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu) + \frac{\pi^2}{\theta_n^2}(w_i w_j + \mu(w_i + w_j) + \mu^2 + \sigma^2) \Big)^{n-2} \Big\}. \qquad (38)
\end{aligned}
$$

Now consider p^2. Using the fact that μ, σ and $1/2 - (\hat{w} + \mu)\pi/\theta_n$ are all $\Theta(1)$,

$$
\begin{aligned}
p^2 &= \int f(w_i)dw_i \int f(w_j)dw_j (1 - \frac{\pi}{\theta_n}(w_i + \mu))^{n-1} (1 - \frac{\pi}{\theta_n}(w_j + \mu))^{n-1} \\
&\geq \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j (1 - \frac{\pi}{\theta_n}(w_i + \mu))^{n-1} (1 - \frac{\pi}{\theta_n}(w_j + \mu))^{n-1} \qquad (39) \\
&= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j \Big(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu) + \frac{\pi^2}{\theta_n^2}(w_i w_j + \mu(w_i + w_j) + \mu^2) \Big)^{n-1} \\
&= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j \Big(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu) + \frac{\pi^2}{\theta_n^2}(w_i w_j + \mu(w_i + w_j) + \mu^2 + \sigma^2) \Big)^{n-1} (1 - o(1)) \\
&> \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j \Big(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu \Big) \times \\
&\quad \Big(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu) + \frac{\pi^2}{\theta_n^2}(w_i w_j + \mu(w_i + w_j) + \mu^2 + \sigma^2) \Big)^{n-2} (1 - o(1)) \\
&= \int_0^{\hat{w}} f(w_i)dw_i \int_0^{\hat{w}} f(w_j)dw_j \Big(1 - \frac{\pi}{\theta_n}(w_i + w_j) \Big) \times \\
&\quad \Big(1 - \frac{\pi}{\theta_n}(w_i + w_j + 2\mu) + \frac{\pi^2}{\theta_n^2}(w_i w_j + \mu(w_i + w_j) + \mu^2 + \sigma^2) \Big)^{n-2} (1 - o(1)). \qquad (40)
\end{aligned}
$$

Finally, this gives the desired ratio

$$
\frac{\Pr[Y_i = 1, Y_j = 1]}{p^2} < 1 + o(1). \qquad (41)
$$

By the second moment method it then follows that $\Pr[Y = 0] < o(1)$. □

5.2 Connected Graph

Theorem 4. *Let $\theta_n = cn/\log n$ for $c < \sup_{\alpha \in (0,1)} \alpha F^{-1}(1-\alpha)/4$. Then the GTG is connected whp.*

Proof. The proof is divided into two parts. In the first part, we prove that a constant fraction of nodes αn are connected. In the second part we prove that the rest of the $(1-\alpha)n$ nodes are connected to the first set of αn nodes.

First part: Invoking the proof of the appearance of the giant component, there are αn nodes all with weights $\geq s_n \to F^{-1}(1-\alpha) = \Theta(1)$.

Let $\theta_n = cn/\log n$, where $c < \alpha F^{-1}(1-\alpha)\pi$. Analogously to r_t, define the connectivity radius r_c

$$r_c^2(w_i) = \frac{w_i}{\theta_n} \geq \frac{s_n}{\theta_n} > \frac{\log n}{\alpha\pi n}\,. \tag{42}$$

Similarly to Theorem 2 let us consider small circles around each of these αn nodes, and consider these nodes as a RGG. It is known that $r_n = \sqrt{\log n/(\pi n)}$ is the connectivity threshold in RGG [GK98]. The connectivity of RGG implies the connectivity of these αn nodes in our GTG.

Second part: Color the αn high-weighted nodes blue, and the remaining $(1-\alpha)n$ nodes red. Now let us tile our space into $n/(c_0 \log n)$ squares of size $c_0 \log n/n$. We state c_0 later. Consider any square S_i, and let B_i be the number of blue nodes in S_i. In expectation there are $\mathrm{E}[B_i] = \alpha c_0 \log n$ blue nodes in each square. From the Chernoff bound it follows

$$\Pr[B_i \geq (1-\delta)\alpha c_0 \log n] \geq 1 - n^{-\alpha c_0 \delta^2/2}. \tag{43}$$

Let us consider one red node r. The node r belongs to some square S_i. Let M_r be the event that the red node r is connected to some blue node $b \in S_i$. Let the weights of r, b be w_r, w_b, respectively. The probability of the complement of M_r, conditioned on there being at least one blue node in S_i, is given by

$$\begin{aligned}\Pr[M_r^c | B_i \geq 1] = \Pr[w_r + w_b \leq r^2\theta_n] &\leq \Pr[w_r + w_b \leq 2c_0 \frac{\log n}{n} c \frac{n}{\log n}] \\ &= \Pr[w_r + w_b \leq 2c_0 c]. \end{aligned}\tag{44}$$

As long as $F^{-1}(1-\alpha) > 2c_0 c$, $w_b > 2c_0 c$ and hence $\Pr[M_r^c | B_i \geq 1] = 0$. For large enough n it must hold that $(1-\delta)\alpha c_0 \log n > 1$, and so from Eq. (43),

$$\begin{aligned}\Pr[M_r^c] &\leq \Pr[M_r^c | B_i \geq (1-\delta)\alpha c_0 \log n] + \Pr[B_i < (1-\delta)\alpha c_0 \log n] \\ &\leq 0 + n^{-\alpha c_0 \delta^2/2}. \end{aligned}\tag{45}$$

If $\alpha c_0 \delta^2/2 \geq 1 + \varepsilon$ for some $\varepsilon > 0$, then by the union bound,

$$\Pr[\bigcup_r M_r^c] \leq \sum_r \Pr[M_r^c] \leq (1-\alpha) n n^{-(1+\varepsilon)} = (1-\alpha) n^{-\varepsilon}. \tag{46}$$

Finally, the probability that all red nodes are connected to the set of blue nodes is given by the following relation

$$\Pr[\bigcap_r M_r] = 1 - \Pr[\bigcup_r M_r^c] \geq 1 - (1-\alpha) n^{-\varepsilon} \to 1. \tag{47}$$

The requirements we have imposed on constants so far are: $c < \alpha F^{-1}(1-\alpha)\pi$, $c < F^{-1}(1-\alpha)/(2c_0)$ and $\alpha c_0 \geq 2(1+\varepsilon)/\delta^2$. These conditions combine to give

$$c < \alpha F^{-1}(1-\alpha) \min(\pi, \frac{\delta^2}{4(1+\varepsilon)}). \tag{48}$$

Since $\alpha \in (0,1)$, $\delta \in (0,1)$ and $\varepsilon > 0$ are arbitrary, we obtain

$$c < \sup_{\alpha\in(0,1)} \alpha F^{-1}(1-\alpha)/4. \tag{49}$$

□

6 Diameter

In this section we analyze the diameter of GTG, and provide an upper bound on it. In the design of large networks, such as the Internet, wireless networks, etc., it is desirable to achieve low latency in the graph (i.e., the hop-count between any pair of nodes in the network is small). In other words, a graph with a small diameter is desired. We give conditions on the threshold function θ_n such that the graph has a desired diameter in general. Furthermore, we derive conditions on θ_n, in terms of the cumulative distribution function on weights $F(w)$, such that *diam* belongs to the specific classes $diam = O(1)$, $diam = O(\log^q n)$ and $diam = O(\sqrt{n/\log n})$. These correspond to an ultra-low, low and high latency network, respectively. For these three classes, we give the exact expressions on θ_n in the case of the exponentially distributed weights. Note that all of these classes correspond to denser graphs than those we have considered so far, i.e., with θ_n scaling as $o(\log n/n)$ vs. the $\Theta(\log n/n)$ scaling for connectivity.

Let u and v be two arbitrary nodes. Construct a sequence of adjacent squares $S_1, S_2, \ldots, S_{O(1/x)}$, of size $x \times x$, linking u and v, such that u and v are the centers of the first and last squares respectively[4] (see Fig. 2). The geometric distance between any two nodes is $r \leq \Theta(1)$. Thus, there are $O(1/x)$ squares on the straight path $u - v$ in total.

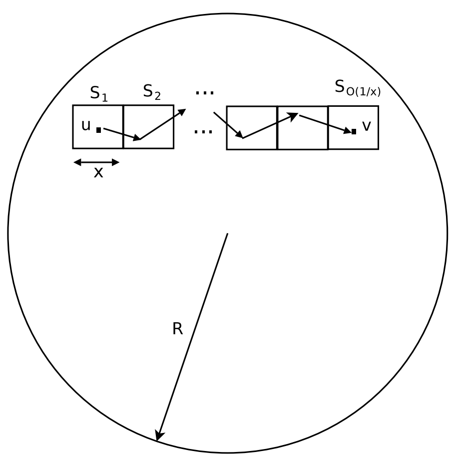

Fig. 2. Illustration of our diameter proof technique: a sequence of adjacent squares of size $x \times x$ link an arbitrary pair of nodes u and v in a unit-area disc.

[4] The centers of the squares lie on the straight line $u - v$.

Let V_i be the number of nodes that lie within the square S_i, for $i = 1, 2, \dots, O(1/x)$. We have $\mathrm{E}[V_i] = nx^2$. Using the Chernoff bound, the following is satisfied

$$\Pr[V_i \le (1-\delta)\mathrm{E}[V_i]] \le e^{-\mathrm{E}[V_i]\delta^2/2}. \tag{50}$$

Taking $\delta = 1/2$, we get $\Pr[V_i \le nx^2/2] \le e^{-nx^2/8}$, i.e., in each square S_i, there are at least $nx^2/2$ nodes whp.

Let M_i be the event that in a square S_i, there is at least one node with weight $w \ge s_n$. We will specify s_n later. Now, we derive a lower bound on the probability $\Pr[M_i]$. This probability is greater than the probability conditioned on the event that there are at least $nx^2/2$ nodes in S_i, i.e.

$$\begin{aligned}\Pr[M_i] &\ge \Pr[M_i | V_i \ge nx^2/2]\Pr[V_i \ge nx^2/2] \\ &\ge (1 - \Pr[W \le s_n]^{nx^2/2})(1 - e^{-nx^2/8}) \\ &= (1 - F(s_n)^{nx^2/2})(1 - e^{-nx^2/8}). \end{aligned} \tag{51}$$

We now explain how we choose s_n such that any two neighboring squares S_j and S_{j+1} are connected by an edge (i.e., there are two connected nodes $a \in S_j$ and $b \in S_{j+1}$). Let weights of a and b be w and w', respectively. We showed that in any square S_i there is at least one node with weight $\ge s_n$, whp. We want that the connectivity relation Eq.(1) for nodes a and b is satisfied. Maximal distance $||a-b||$ between a pair of nodes is $||a-b|| \le x\sqrt{5}$. Conditioned on the events that weights w, w' are greater then s_n we have the following relation for the connectivity of nodes a and b

$$\Pr[a \sim b | w, w' \ge s_n] \ge \Pr[2s_n/r^2 \ge \theta_n]. \tag{52}$$

Let us choose $s_n = \Theta(x^2\theta_n)$. If an arbitrary pair of nodes (u, v) is connected by a path of nodes belonging to the squares $S_1, S_2, \dots, S_{O(1/x)}$, then the following relation on *diam* is satisfied

$$\begin{aligned}\Pr[diam = O(1/x)] &\ge \Pr[\cap_{i=1}^{O(1/x)} M_i] \\ &= \left((1 - e^{-nx^2/8})(1 - F(s_n)^{nx^2/2}\right)^{O(1/x)}, \end{aligned}$$

since the nodes, as well as weights, are distributed independently. Now, the lemma on the diameter follows.

Lemma 1. *Let the cumulative weight distribution function be $F(w)$ in GTG model. Let x and a sequence $s_n = \Theta(x^2\theta_n)$ be such that*

$$\lim_{n\to\infty} \left(1 - F(s_n)^{nx^2/2}\right)^{1/x} = 1. \tag{53}$$

Then, whp $diam = O(1/x)$.

Proof. Proof follows from the previous discussion.

6.1 Some Classes of Diameter

We now analyze and state conditions on θ_n such that $diam = O(1)$, $diam = O(\log^q n)$ and $diam = O(\sqrt{n/\log n})$. We work out the case when the weight distribution is exponential, $f(w) = e^{-w}, w \geq 0$, (i.e., $F(w) = 1 - e^{-w}, w \geq 0$) and derive an upper bound on the threshold function θ_n in this particular case. For some other weight distribution, the analysis would be similar. Our results are given by:

1. **Ultra-low Latency:** $diam = O(1)$**.** Let $x < 1$ be a constant and $s_n = \theta_n$. If $F(\theta_n)^n \to 0$, then $diam = O(1)$ whp. For the exponential weight distribution it follows that $\theta_n = o(\log n)$.
2. **Low Latency:** $diam = O(\log^q n)$**.** Let $x = 1/\log^q n$ and $s_n = \theta_n/\log^{2q} n$. If $F(\theta_n/\log^{2q} n)^{\frac{n}{2\log^{2q} n}} \log^q n \to 0$, then $diam = O(\log^q n)$ whp. For the exponential weight distribution it follows that $\theta_n = o\Big((\log n)^{2q(1-(\log^{2q} n)/n)}\Big)$.
3. **High Latency:** $diam = O(\sqrt{n/\log n})$**.** Let $x = \sqrt{\log n/n}$ and $s_n = \theta_n \log n/n$. If $\sqrt{n/\log n} F(\theta_n \log n/n)^{\log n} \to 0$, then $diam = O(\sqrt{n/\log n})$ whp. For the exponential weight distribution it follows that $\theta_n = o\Big((n/\log n)^{1-1/(2\log n)}\Big)$.

Here we prove the previous claims.

Ultra-low Latency: $diam = O(1)$**.** For the diameter to be a constant, let $x < 1$ be a constant. Invoking Lemma 1, it follows that $diam = O(1)$ whp if and only if $1 - F(s_n)^{nx^2/2} \to 1$, i.e., if and only if $F(s_n)^n \to 0$. The condition on the size of $diam$ is given by the following claim, and we can derive θ_n such that $diam = O(1)$ whp.

Claim. If $F(\theta_n)^n \to 0$, then $diam = O(1)$ whp.

For the exponential weight distribution it follows that $F(\theta_n)^n = (1 - e^{-\theta_n})^n \to e^{-n/e^{\theta_n}}$. The last equation tends to 0 if and only if $n/e^{\theta_n} \to \infty$. That is,

Claim. For the exponential weight distribution $f(w) = e^{-w}$, the diameter in GTG is $diam = O(1)$ if $\theta_n = o(\log n)$.

Low Latency: $diam = O(\log^q n)$**.** Let us choose $x = 1/\log^q n$. Invoking Lemma 1, we obtain:

$$(1 - F(s_n)^{nx^2/2})^{1/x} = \left(1 - F(s_n)^{\frac{n}{2\log^{2q} n}}\right)^{\log^q n} \tag{54}$$

For $s_n \to 0$, the last expression tends to 1, if and only if

$$F(s_n)^{\frac{n}{2\log^{2q} n}} \log^q n \to 0, \tag{55}$$

by using $\lim_{t\to+\infty}(1-1/t)^t = 1/e$. The condition on the size of $diam$ is given by the following claim.

Claim. if $F(\theta_n/\log^{2q} n)^{\frac{n}{2\log^{2q} n}} \log^q n \to 0$, then $diam = O(\log^q n)$ whp.

For the exponential weight distribution, the following is to be satisfied

$$F(s_n)^{\frac{n}{2\log^{2q} n}} \log^q n = \log^q n (1 - e^{-s_n})^{\frac{n}{2\log^{2q} n}} \to s_n^{n/(2\log^{2q} n)} \log^q n \to 0, \tag{56}$$

or equivalently

$$s_n = o\Big((\log n)^{-\frac{2q}{n}\log^{2q} n)}\Big). \tag{57}$$

Claim. For the exponential weight distribution $f(w) = e^{-w}$, the diameter in GTG is $diam = O(\log^q n)$ if $\theta_n = o\Big((\log n)^{2q(1-(\log^{2q} n)/n)}\Big)$.

High Latency: $diam = O(\sqrt{n/\log n})$**.** Let us choose $x = \sqrt{\log n/n}$. Invoking Lemma 1, we get

$$(1 - F(s_n)^{nx^2/2})^{1/x} = (1 - F(s_n)^{\log n})^{\sqrt{\frac{n}{\log n}}} \tag{58}$$

It can be shown that the last expression tends to 1 if and only if $\sqrt{n/\log n}F(s_n)^{\log n} \to 0$, by using $\lim_{t\to+\infty}(1 - 1/t)^t = 1/e$. The condition on the size of *diam* is given by the following claim.

Claim. If $\sqrt{n/\log n}F(\theta_n \log n/n)^{\log n} \to 0$, then $diam = O(\sqrt{n/\log n})$ whp.

For the exponential weight distribution, the following is to be satisfied

$$\sqrt{n/\log n}F(s_n)^{\log n} = \sqrt{n/\log n}(1 - e^{-s_n})^{\log n} \to \sqrt{n/\log n}s_n{}^{\log n} \to 0, \tag{59}$$

or equivalently

$$s_n = o\Big((\log n/n)^{1/(2\log n)}\Big). \tag{60}$$

Claim. For the exponential weight distribution $f(w) = e^{-w}$, the diameter in GTG is $diam = O(\sqrt{n/\log n})$ if $\theta_n = o\Big((n/\log n)^{1-1/(2\log n)}\Big)$.

Simulation results are shown for the GTG with path-loss exponent $\beta = 3$ (not $\beta = 2$), for the case of $diam = O(\log^{1.5} n)$ in Fig. 3 and $diam = O(\sqrt{n/\log n})$ in Fig. 4. Exponentially distributed weights with mean 1 are used. The network sizes simulated are $n = \{100, 200, 500, 1000, 2000, 10000\}$. The threshold values θ_n for the two cases are obtained by invoking previous claims.

7 Clustering Coefficient

7.1 Weights Given

Let us consider in more detail the clustering coefficient defined in Section 5, namely the neighbor probability

$$C(w_i, w_j, w_k) = \Pr[v_i \sim v_j | v_i \sim v_k, v_j \sim v_k, w_i, w_j, w_k]. \tag{61}$$

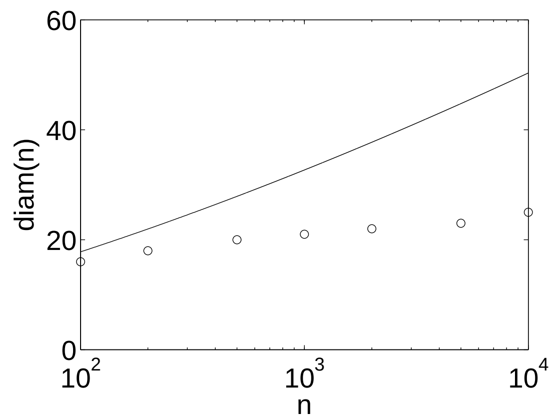

Fig. 3. (a) For the case of $diam = O(\log^q n)$, with $q = 1.5$, the analytical solid curve is the upper bound on $diam(n)$. Simulation results match with theoretical predictions, since the simulation points all lie below the analytical curve.

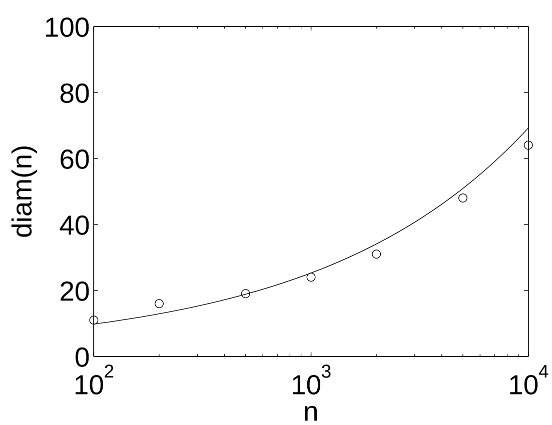

Fig. 4. (b) For the case of $diam = O(\sqrt{n/\log n})$, the solid curve plots the upper bound on $diam(n)$, and this bound closely matches the experimental values.

Let $x = \sqrt{(w_i+w_j)/\theta_n}$, $y = \sqrt{(w_j+w_k)/\theta_n}$ and $z = \sqrt{(w_i+w_k)/\theta_n}$. Then, if d_{ij} represents the distance between points i and j,

$$\begin{aligned} C(w_i,w_j,w_k) &= \Pr[d_{ij} \le x | d_{jk} \le y, d_{ik} \le z] \\ &= \frac{1}{\pi z^2}\int_0^z \Pr[d_{ij} \le x | d_{jk} \le y, d_{ik} = r]\, 2\pi r\, dr \\ &= \frac{1}{\pi^2 y^2 z^2}\int_0^z A(r) 2\pi r\, dr \end{aligned} \tag{62}$$

where $A(r)$ is the overlap area of a disc of radius x, centered at i, and a disc of radius y, centered at k. Now consider a triangle $\triangle ABC$, with sides $AB = r$, $AC = x$, $BC = y$, $\angle CAB = \alpha$, $\angle ABC = \beta$. Following arguments similar to those in [DC02], there are three possible cases for $A(r)$:

$$A(r) = \begin{cases} \pi[\min(x,y)]^2 & r \le |x-y| \\ x^2(\alpha - \sin\alpha\cos\alpha) + y^2(\beta - \sin\beta\cos\beta) & |x-y| < r < x+y \\ 0 & r \ge x+y \end{cases} \tag{63}$$

where

$$\alpha = \cos^{-1}\left(\frac{r^2+x^2-y^2}{2xr}\right), \qquad \beta = \cos^{-1}\left(\frac{r^2-x^2+y^2}{2yr}\right). \tag{64}$$

From the definitions of x, y and z, $|x-y| < z < x+y$. After some algebraic manipulation, one then finds

$$\begin{aligned} C = \frac{x^2}{\pi}\Bigg[& \frac{1}{y^2}\cos^{-1}\left(\frac{z^2+x^2-y^2}{2xz}\right) + \frac{1}{x^2}\cos^{-1}\left(\frac{z^2-x^2+y^2}{2yz}\right) + \frac{1}{z^2}\cos^{-1}\left(\frac{x^2+y^2-z^2}{2xy}\right) - \\ & \frac{x^2+y^2+z^2}{4x^2y^2z^2}\sqrt{2x^2y^2+2x^2z^2+2y^2z^2-x^4-y^4-z^4}\Bigg] \\ = {} & \frac{1}{\pi(w_i+w_k)(w_j+w_k)}\Bigg\{(w_i+w_j)(w_i+w_k)\cos^{-1}\left(\frac{w_i}{\sqrt{w_i+w_j}\sqrt{w_i+w_k}}\right) + \\ & (w_i+w_j)(w_j+w_k)\cos^{-1}\left(\frac{w_j}{\sqrt{w_i+w_j}\sqrt{w_j+w_k}}\right) + \\ & (w_i+w_k)(w_j+w_k)\cos^{-1}\left(\frac{w_k}{\sqrt{w_i+w_k}\sqrt{w_i+w_k}}\right) - \\ & (w_i+w_j+w_k)\sqrt{w_iw_j+w_jw_k+w_kw_i}\Bigg\}. \end{aligned}$$

Note that while C is a function of the weights w_i, w_j and w_k, it is independent of θ_n. This reflects a similar property in random geometric graphs [DC02], where the clustering coefficient is independent of the graph's mean degree.

Written in terms of the connection probabilities a, b and c defined in Theorem 3, C is given by

$$\begin{aligned} C\pi bc = {} & ab\cos^{-1}\frac{a+b-c}{2\sqrt{ab}} + bc\cos^{-1}\frac{-a+b+c}{2\sqrt{bc}} + ca\cos^{-1}\frac{a-b+c}{2\sqrt{ac}} - \\ & \frac{a+b+c}{4}\sqrt{2ab+2ac+2ca-a^2-b^2-c^2}. \end{aligned} \tag{65}$$

We now prove the bound on the clustering coefficient that we needed for Theorem 3.

Lemma 2. *If* $w_i, w_j, w_k \leq \hat{w} = (1 - 3\sqrt{3}/4\pi)\theta_n/2\pi$ *then* $C \geq a$.

Proof. Define

$$\begin{aligned} S(a,b,c) &= \frac{C\pi}{a} \\ &= \frac{1}{a}\cos^{-1}\frac{-a+b+c}{2\sqrt{bc}} + \frac{1}{b}\cos^{-1}\frac{a-b+c}{2\sqrt{ac}} + \frac{1}{c}\cos^{-1}\frac{a+b-c}{2\sqrt{ab}} - \gamma\frac{a+b+c}{4abc}, \end{aligned}$$

where

$$\begin{aligned} \gamma &= \sqrt{2ab+2ac+2ca-a^2-b^2-c^2} \\ &= 2\sqrt{bc}\sqrt{1-\Big(\frac{-a+b+c}{2\sqrt{bc}}\Big)^2} \end{aligned} \tag{66}$$

It is easy to verify that $S = \pi$ when $w_i = w_j = w_k = \hat{w}$. We will now show that S is nonincreasing over the weights, and thus $S \geq \pi$ for all smaller values of w_i, w_j, w_k.

Consider the sign of the derivative

$$\begin{aligned} \frac{dS}{dw_i} &= \frac{\partial S}{\partial a}\frac{\partial a}{\partial w_i} + \frac{\partial S}{\partial b}\frac{\partial b}{\partial w_i} + \frac{\partial S}{\partial c}\frac{\partial c}{\partial w_i} \\ &= \frac{\pi}{\theta_n}\Big(\frac{\partial S}{\partial a} + \frac{\partial S}{\partial b}\Big). \end{aligned} \tag{67}$$

Since S is symmetric in a and b, it is sufficient to consider the sign of $\partial S/\partial a$:

$$\frac{\partial S}{\partial a} = \frac{\gamma}{4a^2bc}(-a+b+c) - \frac{1}{a^2}\cos^{-1}\frac{-a+b+c}{2\sqrt{bc}}. \tag{68}$$

Now let

$$t = \frac{-a+b+c}{2\sqrt{bc}} \in [0,1]. \tag{69}$$

Then,

$$\frac{\partial S}{\partial a} = \frac{1}{a^2}(t\sqrt{1-t^2} - \cos^{-1} t). \tag{70}$$

Given the function $\varphi(t) = t\sqrt{1-t^2} - \cos^{-1} t$ on $[0,1]$, $\varphi'(t) = 2\sqrt{1-t^2} \geq 0$ and $\varphi(1) = 0$. It follows that $\partial S/\partial a \leq 0$, and so $dS/dw_i \leq 0$. Finally, S is symmetric in (w_i, w_j, w_k), so it must be nonincreasing over each of the weights and bounded below by the value at $w_i = w_j = w_k = \hat{w}$. □

7.2 Degree Given

Define C_l to be the neighbor probability for a node with a given degree

$$\begin{aligned} C_l &= \Pr[v_i \sim v_j | v_i \sim v_k, v_j \sim v_k, d(v_k) = l] \\ &= \frac{\Pr[v_i \sim v_j, v_i \sim v_k, v_j \sim v_k, d(v_k) = l]}{\Pr[v_i \sim v_k, v_j \sim v_k, d(v_k) = l]} \\ &= \frac{\int f(\mathbf{w})\Pr[v_i \sim v_j, v_i \sim v_k, v_j \sim v_k, d(v_k) = l|\mathbf{w}]d\mathbf{w}}{\int f(\mathbf{w})\Pr[v_i \sim v_k, v_j \sim v_k, d(v_k) = l|\mathbf{w}]d\mathbf{w}}. \end{aligned}$$

A straightforward calculation shows that C_l is given by the ratio I_n/I_d of two integrals, where

$$I_d = \int f(w_k)\Big(\frac{\pi}{\theta_n}(\mu+w_k)\Big)^l\Big(1-\frac{\pi}{\theta_n}(\mu+w_k)\Big)^{n-l}dw_k \tag{71}$$

and

$$\begin{aligned} I_n &= \int\int\int f(w_i)f(w_j)f(w_k)\pi y^2\pi z^2 C(w_i,w_j,w_k)dw_i dw_j \\ &\quad\times\Big(\frac{\pi}{\theta_n}(\mu+w_k)\Big)^{l-2}\Big(1-\frac{\pi}{\theta_n}(\mu+w_k)\Big)^{n-l}dw_k \\ &= \int\int\int f(w_i)f(w_j)f(w_k)(w_i+w_k)(w_j+w_k)C(w_i,w_j,w_k)dw_i dw_j \\ &\quad\times\frac{1}{(\mu+w_k)^2}\Big(\frac{\pi}{\theta_n}(\mu+w_k)\Big)^{l}\Big(1-\frac{\pi}{\theta_n}(\mu+w_k)\Big)^{n-l}dw_k \end{aligned} \tag{72}$$

For a specific weight distribution, these integrals may be evaluated numerically. It is intuitive that when l is very large ($l=\Theta(n)$), C_l should scale as $1/n$: nodes that connect to very many neighbors presumably do so because of their high weights, and their neighbors are no more likely to be connected than any two random nodes are. Interestingly, in the case shown in Fig. 5, C_l scales almost perfectly as l^{-1} even at relatively small values of l. For an exponential weight distribution with mean 1 and parameters $n=1000$ and $\theta_n=1000$, the slope on the log-log plot already appears very close to -1 at $l=8$.

8 Summary

Geographical threshold graphs are a rich model with the possibility of controlling structural properties by choosing specific weight distributions and tuning threshold values. The model is a versatile one and can be used not only for the generation and analysis of web-graphs or large complex networks, but more generally for relation graphs in a large data set. If the data have a metric and can be mapped to nodes in Euclidean space, much of the foregoing analysis applies: one may hope to control structural properties of the data set by studying it as a GTG.

In this paper we have analyzed some of the structural properties of a GTG. Given a node weight distribution $f(w)$ and threshold θ_n, the degree distribution can be easily calculated. We have given bounds on the threshold value θ_n guaranteeing the absence and existence of the giant component. We have also given bounds on θ_n guaranteeing a disconnected and connected graph, and provided upper bounds on the diameter for sufficiently dense graphs. Finally, we have derived a formula for the clustering coefficient in terms of the weight distribution and threshold, as well as discussed the general clustering coefficient for nodes with a given degree, and evaluated it numerically. Our analysis has used the additive threshold function $(w_i+w_j)/r^2 \geq \theta_n$ for the connectivity relation, but not all of our techniques require it. For this reason, many of the results may be generalized to other threshold functions, other path-loss exponents and other spatial dimensionality in a straightforward manner.

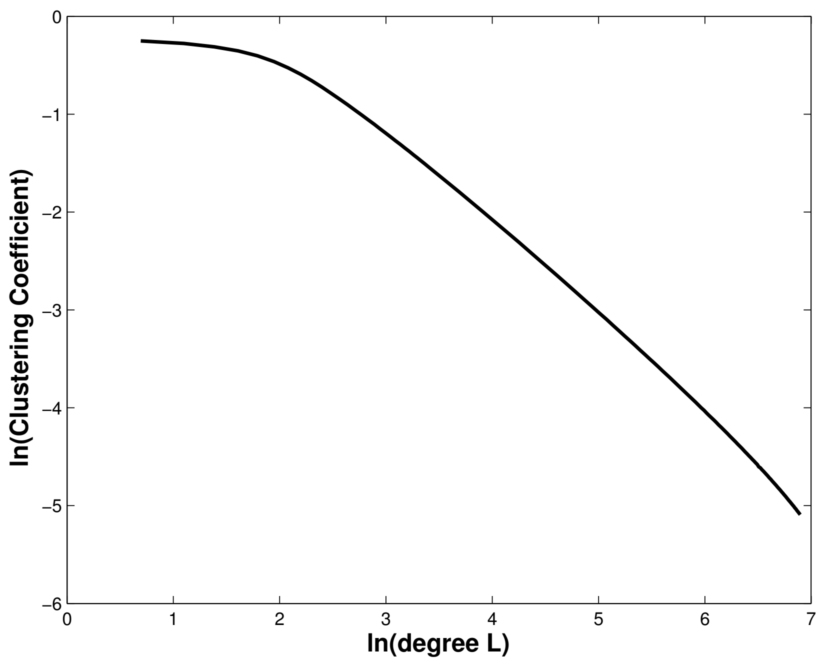

Fig. 5. Clustering Coefficient vs. Degree L

Acknowledgements

Part of this work was funded by the Department of Energy at Los Alamos National Laboratory under contract DE-AC52-06NA25396 through the Laboratory Directed Research and Development Program.

References

[ACL00] William Aiello, Fan Chung, and Linyuan Lu. A random graph model for massive graphs. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 171–180, New York, NY, USA, 2000. ACM Press.

[APR02] James Abello, Panos M. Pardalos, and Mauricio G. C. Resende, editors. *Handbook of massive data sets*. Kluwer Academic Publishers, Norwell, MA, USA, 2002.

[AS00] Noga Alon and Joel H. Spencer. *The probabilistic method*. John Wiley & Sons, Inc., New York, 2nd edition, 2000.

[BA99] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509 – 512, 1999.

[BHP07] Milan Bradonjić, Aric A. Hagberg, and Allon G. Percus. Giant component and connectivity in geographical threshold graphs. In Anthony Bonato and Fan R. K. Chung, editors, *Proceedings of the 5th Workshop on Algorithms and Models for the Web-Graph (WAW2007)*, volume 4863 of *Lecture Notes in Computer Science*, pages 209–216. Springer, 2007.

[BK07] Milan Bradonjić and Joseph Kong. Wireless ad hoc networks with tunable topology. In *Proceedings of the 45th Annual Allerton Conference on Communication, Control and Computing*, 2007.

[Bol01] Béla Bollobás. *Random Graphs*. Cambridge University Press, 2nd edition, 2001.

[Bon05] Anthony Bonato. A survey of models of the web graph. In *Combinatorial and Algorithm Aspects of Networking*, volume 3405 of *Lecture Notes in Computer Sciences*, pages 159–172, Heidelberg, 2005. Springer Berlin.

[BRST01] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Struct. Algorithms*, 18(3):279–290, 2001.

[CF01] Colin Cooper and Alan M. Frieze. A general model of undirected Web graphs. In *ESA '01: Proceedings of the 9th Annual European Symposium on Algorithms*, pages 500–511, London, UK, 2001. Springer-Verlag.

[CL06] Fan Chung and Linyuan Lu. The volume of the giant component of a random graph with given expected degrees. *SIAM J. Discret. Math.*, 20(2):395–411, 2006.

[DC02] Jesper Dall and Michael Christensen. Random geometric graphs. *Physical Review E*, 66:016121, 2002.

[Dur06] Rick Durrett. *Random Graph Dynamics*. Cambridge University Press, 1st edition, 2006.

[ER59] P. Erdős and A. Rényi. On random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 1959.

[ER60] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 1960.

[FFF99] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pages 251–262, New York, NY, USA, 1999. ACM Press.

[GK98] P. Gupta and P. R. Kumar. Critical power for asymptotic connectivity. In *Proceedings of the 37th IEEE Conference on Decision and Control*, volume 1, pages 1106–1110, 1998.

[HSS06] Aric Hagberg, Pieter J. Swart, and Daniel A. Schult. Designing threshold networks with given structural and dynamical properties. *Phys. Rev. E*, 74:056116, 2006.

[KKR+99] Jon M. Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S. Tomkins. The web as a graph: Measurements, models, and methods. In *Computing and Combinatorics: 5th Annual International Conference, COCOON'99, Tokyo, Japan, July 1999. Proceedings*, volume 1627, pages 1–17, Heidelberg, 1999. Springer Berlin.

[KRR+00] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *FOCS '00: Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, page 57, Washington, DC, USA, 2000. IEEE Computer Society.

[KY06] Zhenning Kong and Edmund M. Yeh. Analytical lower bounds on the critical density in continuum percolation. arXiv:math/0610751, 2006.

[MMK05] Naoki Masuda, Hiroyoshi Miwa, and Norio Konno. Geographical threshold graphs with small-world and scale-free properties. *Physical Review E*, 71(3):036108, 2005.

[MP95] N. V. R. Mahadev and U. N. Peled. *Threshold graphs and related topics*, volume 56 of *Annals of discrete mathematics*. Elsevier, New York, 1995.

[MR95] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence, 1995.

[Pen03] Mathew D. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.